IMPROVING BACKUP PROCESS QUALITY AND RESILIENCE USING HYBRID CLOUD INFRASTRUCTURES

Marius Ioan TODERICI
National University of Science and Technology Politehnica Bucharest
Bucharest, Romania,
marius@toderici.ro

Aurel Mihail TITU*
"Lucian Blaga" University of Sibiu
Sibiu, Romania,
mihail.titu@ulbsibiu.ro

Oana Roxana CHIVU
National University of Science and Technology Politehnica Bucharest
Bucharest, Romania,
oana.chivu@upb.ro

ABSRACT
Backup processes are vital for any organization, they ensure data recovery in case of accidental data loss and most importantly they ensure data recovery in case of a disaster or a cyber-attack. From this point of view backup processes are very important they are essential to ensure the resilience of the organization. The existence of well-established backup processes, regular testing of these processes is very important for any organization so that the implementation of these processes ensures rapid and secure recovery of data in the event of data loss but also ensures the resilience of the organization in the event of a disaster. Storage capacity is one of the most expensive resources available in the cloud in the long term when the amount of data is very large. In the case of hybrid cloud systems, a thorough analysis of the data that is stored is required to determine where this data is processed, how often the data is accessed, the medium and long-term storage costs, the bandwidth required for traffic and the time required for transfer. Cloud storage has several advantages such as access to data from anywhere, ensuring a data recovery environment in case of disaster, scalability and security. Large and very large enterprise organizations produce a very large amount of data that must be available from anywhere, provide scalability, ensure continuity in case of disaster, allow data control and be cost-effective. Storing data in the public cloud also has the advantage of lower maintenance costs but these will be offset, as we have mentioned, by the communications infrastructure that must ensure sufficient bandwidth for data traffic. For these reasons, hybrid cloud is a very good solution because it takes advantage of both cloud and local resources by combining public and private cloud capabilities, forming an integrated storage architecture.

Keywords: capacity management, cloud storage, 3-2-1-1 backup, management processes

1. INTRODUCTION

Hybrid cloud infrastructure is a combination in which on-premises IT infrastructure is interfaced with cloud systems and work together to achieve the organization's IT goals to support the organization's processes. The on-premises infrastructure generally acts as a private cloud, processes that are data-intensive and bandwidth-intensive are often hosted here. To ensure compatibility of systems the hybrid infrastructure uses a common software platform on the supervision side. This platform is intended to enable the management of compute, storage, storage, etc. resources in the hybrid IT infrastructure. If one were to define hybrid cloud infrastructure (Figure 1) it is a collection of cloud-based and on-premises IT resources that are

interconnected and work together to provide high availability of systems, increased agility and flexibility that enable much faster adaptation to the workloads of the digital age.

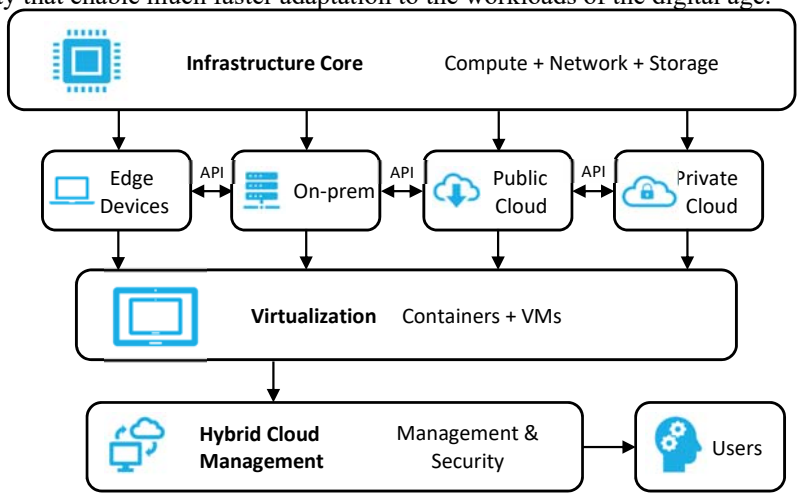

Figure 1. Hybrid cloud infrastructure

The hybrid cloud architecture consists of cloud-based Infrastructure-as-a-service (IaaS) platforms from different vendors such as Amazon Web Services, Microsoft Azure, Google Cloud or VMware Cloud. The IaaS infrastructure is interconnected with the on-premises infrastructure using broadband communication systems in order to be able to handle the communication needs and the high volume of data exchange within this ecosystem.
The main advantages of hybrid infrastructure are given by the flexibility to host the organization's systems and applications. In the case of legacy systems or legislator-regulated systems that cannot be hosted on public systems, they will be hosted on on-premises private cloud infrastructure, while systems that need distributed access or seasonal resources will be hosted in the cloud. The scalability and agility of the hybrid infrastructure ensures that IT resources are needed when they are needed and can be reduced when they are not, ensuring economic efficiency.

2. BACKUP METHODS AND PROCEDURES

Regardless of the storage medium used for the data being backed up, IT administrators need to ensure redundant capabilities of the equipment by using RAID arrays or backing up data in two or more locations simultaneously so that access can be assured in the event that some of the storage components fail. Ensuring data access being one of the key components of the data storage environment. IT management must ensure procedures for backing up the data of key applications and services.
The resilience of the organization’s IT environment and ensuring the continuity of the organization are very important factors that play a determining role in the organization’s IT management. Analyzing the storage capacities is important to ensure their backup on different storage media that can be used for disaster recovery operations in case of disaster.
When using a hybrid cloud, we need a few components in order to backup efficiently, quickly and securely [1]:

- A storage system that is available on premises and has sufficient capacity to back up systems. The storage system must ensure high data availability (have RAID arrays for storage).
- On the cloud side, the required capacity to be procured, the estimated length of time that cloud storage is required and the possibility to expand this capacity when needed must be calculated.

- On the networking side both locally and externally the bandwidth required for high data traffic will be considered. Equipment must be sized to allow data traffic without disrupting operational activities. Ideally, separate networks and connections should be provided for this activity.
- Use a backup and restore solution compatible with hybrid cloud as this offers flexibility, efficiency and cost savings.

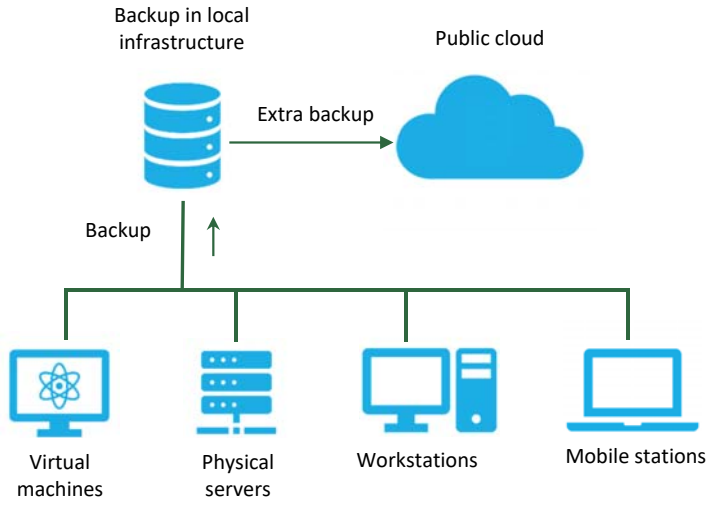

Figure 2. Backup model in hybrid cloud infrastructure (Source nakivo.com)

#

Effective hybrid cloud backup requires training of IT staff so that backup and restore operations can be carried out efficiently and properly with data security and protection in mind (**Error! Reference source not found.**). The learning process and regular exercises require time, effort and resources but will bring the necessary expertise and knowledge to the IT team. At the same time, regular exercises and testing of systems on the backup and restore side will enable the creation and improvement of backup and restore procedures necessary to effectively recover data when needed.

With the hybrid cloud at your disposal the first method of local protection that is also simple to implement is to provide backup to local systems. This method is also the fastest on both the backup and recovery side. Subsequently, data backup will be transferred to the cloud or to another location to minimize data loss if that location is inaccessible. On the cloud side all systems in the cloud will be backed up to the cloud, ideally to other locations as most cloud service providers have this policy in place, and then local backups will be made to their own systems.

What is important to note is that providing backup and restore processes in hybrid cloud infrastructure offers the advantages of storing backups in both on-premises storage and cloud systems and at the same time neutralizes their weaknesses. The advantages are the possibility to ensure short backup windows and fast recovery with minimal downtime. The only disadvantage may be the increased complexity of the infrastructure, but this can be addressed by training and expertise of IT staff, creating clear backup and restore procedures and periodically testing them on test environments.

One of the best and most reliable backup strategies is the 3-2-1-1 strategy, which helps to increase an organization's resilience against ransomware attacks or other disasters.

3. THE 3-2-1-1 BACKUP METHOD

Creating a data backup is no longer sufficient for restoring data in case applications are unable to function or data is lost for various reasons whether it is a security breach, a problem in the infrastructure or other reasons beyond the control of the IT environment. In these situations, the 3-2-1 backup method is indicated. This method is based on having 3 copies of the data, of which 2 copies will be stored on different storage media and one copy will be saved in a location outside the organization's active environment.

The 3 copies of data are the production environment containing the primary data and two backup copies. In practice it has been proven that providing 3 copies of data minimizes the risk of data loss and it can be recovered and restored in any disaster scenario. Having backups in the same location or on the same storage medium increases the likelihood that in the event of a fault or failure both the productive environment and the backups will be unavailable. From this point of view, it is important to back up backups on different storage media. Following the same idea, it is important that the location in which a backup is stored is different from the other as this minimizes the risk that in the event of natural disasters or accidents such as fires, prolonged power failures, it is possible to ensure that data can be recovered from a location that is not altered or blocked for the same reason.

This method can be enhanced by adding an additional copy of the 3-2-1-1 (Figure 3) data to a backup solution that only allows data to be read once the backup data has been saved, it does not allow data to be written or accessed by digital systems. This helps to protect data against ransomware attacks which are increasingly common these days and are favored by attackers because they completely damage victims' data.

Figure 3. The 3-2-1-1 Backup Method

#

For this 3-2-1-1 backup method [2], the hybrid cloud is the conducive environment for this scenario because one of the most common practices is to keep a backup in the on-premises infrastructure and another backup will be kept in the cloud if the productive environment is hosted in the on-premises infrastructure. If the productive environment is in the cloud, then one backup will be secured in another cloud location and one in the organization's local infrastructure. To protect the organization from ransomware attacks and to increase resiliency a copy must be offline or immutable.

4. DATA RETENTION POLICY

Data retention policy is a very important element of capacity management and is the starting point for implementation. This policy defines the data that an organization retains for operational or compliance needs. It describes why the organization needs to retain the data, the retention period and the disposal process. This policy helps an organization determine how to search and access data. Retention policy is particularly important because an organization is continuously producing and consuming data and as the volume of data continually increases this policy is necessary to reduce storage space and costs. Without this data management policy, an organization tends to store data for very large, indefinite periods of time leading to high operating costs, oversizing of storage capacities and storage equipment which translates into increased infrastructure costs, increased backup operations, etc. [3]

In organisations that do not have a data storage and retention policy, IT management needs to make a calculation of the volume of data that is generated on a daily basis, on the data processing side, and where it is used in the cloud or locally. Data that is used daily in production processes, financial data used in processing and reporting internally or to the authorities is more amenable to local storage on fast storage systems and data from customer records is easier to manage in the cloud. Requirements for fast data access is another factor. If we have applications that use real-time data or processes that need fast low-latency access to data or large volumes of data then that data will be stored locally or in the private cloud to minimize access times and avoid potential bottlenecks or data loss caused by inability to access data in a timely manner. For manufacturing applications or industrial or instrumented environments such as SCADA systems, local storage is suitable. But if you have dispersed consumers then cloud storage is an option worth considering.

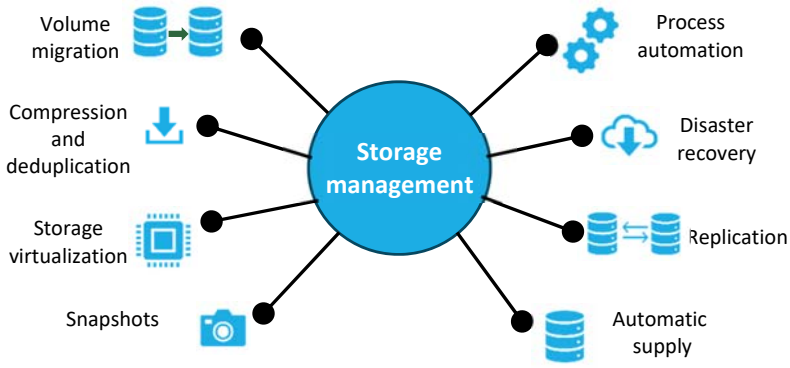

Figure 4. Storage management

On the storage media and storage capacity analysis side, IT management needs to make an analysis of the volume of data being generated and stored daily, monthly and annually on the organization's main systems and services. This analysis is important for estimating the medium- and long-term storage space in both in-house infrastructure and public cloud. Here, managers or administrators need to take into account both the organization's business systems such as production systems, email, file storage, databases as well as IT systems logs.

In the case of organisations using a hybrid cloud, the data storage strategy needs to take into account the data storage policy and data retention period so that it integrates current applications and services as well as future needs. Consider the overheads costs of on-premises and cloud equipment. For storage equipment located in the private cloud, consideration will be given to storage capacity on scalable capacity environments that allow additional space to be added as the organization needs it.

5. COST MANAGEMENT AND RESOURCE OPTIMIZATION

Cost management and resource optimization are necessary to ensure efficient use of storage resources whether storage systems are on-premises in the hybrid cloud or in the cloud.

On the storage side we have identified several cost management and resource optimization methods. From analyzing data from the main systems used in organisations there are some common patterns that are worth addressing in order to reduce storage costs and make efficient use of space whether we are talking about on cloud or local storage. The messaging system (email) is one of the basic pillars of any organization, email messages are a major part of network traffic and a serious contender for the number 1 spot on the storage side. We send a lot of emails with attachments to different people in the organization who in turn send the same email with attachments with possibly little text added so in order not to store the same type of

attachment n times there are systems that identify the attachments index them by ensuring they are stored once and then the system will reference that attachment. This method is known as deduplication and the big advantage is that it allows a file to be stored once no matter how many email addresses it is referenced by.
On the file storage side, collaborative working systems or scanning systems that use the same file storage principle can use the deduplication system, considerably reducing the amount of information that is stored at the end and bringing a speed boost to the infrastructure because once the information is indexed, the system works with links and references that are insignificant from a data storage point of view even if this information is circulated to many users. All modern storage systems have this capability which greatly reduces the amount of space taken up in storage systems by up to 25:1 or more.
Another way to make storage space more efficient is to use back-up archiving and data encryption. This process reduces the amount of data that is transferred over the network or to the cloud bringing serious cost savings when utilized correctly, and this can be achieved through backup policy. All data that is transmitted to the cloud can be archived to ensure the best possible compression rate which translates into lower transfer times and less storage space utilized.
Another method of optimizing long-term storage costs is to periodically filter out backups that are not important. This can be done through the backup policy, periodically test backups held in the cloud for data consistency and delete older copies. Eventually if required for historical purposes, data will be stored on low-cost tape systems which will be stored in at least two different locations. On the cloud storage side, it is very important the period of time over which the data will be stored, the space used as well as the transfer speed and for enterprise organisations that need large storage space it is important to contract for longer periods of time to ensure the lowest cost/megabyte and time. [4]
IT management must analyze the main data to be stored in the cloud and the period of time over which this data will be stored, taking into account the costs but also ensuring the availability of systems and their level of criticality, as it is known that ensuring availability as close to 99.99% is very expensive.

6. CONCLUSIONS

Large organizations today produce large amounts of data that must be managed, analyzed and processed. At the same time, this data must be secured through backup processes that ensure the necessary retention to ensure the organization's resilience in the event of attacks or disasters. From these points of view, the use of a hybrid cloud infrastructure is necessary because it has the ability to efficiently combine the use of the latest technologies on the private cloud side with the versatility of the public cloud. At the same time, by managing the two types of infrastructure, it is much easier to ensure continuity in the event of a disaster, which ensures the resilience of the organization.

7. REFERENCES

[1] Raut, S.; Patil, S.; Hsu, V.: Efficient Backup Management in Hybrid Cloud Deployment Based on Workload Data Classification, 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), Bangalore, India, 2024, pp. 297-300
[2] Childerhose, G.: Mastering Veeam Backup & Replication: Secure backup with Veeam 11 for defending your data and accelerating your data protection strategy , Packt Publishing, 2022.
[3] Li, J.; Singhal, S.; Swaminathan, R.; Karp, A. H.: Managing Data Retention Policies at Scale, IEEE Transactions on Network and Service Management, vol. 9, no. 4, pp.393-406, December 2012
[4] Sok, S.; Plewnia, C.; Tanachutiwat S.; Lichter, H.: Optimization of Compute Costs in Hybrid Clouds with Full Rescheduling, 2020 IEEE International Conference on Smart Cloud (SmartCloud), Washington, DC, USA, 2020, pp. 35-40.